



**CHALMERS**  
UNIVERSITY OF TECHNOLOGY



UNIVERSITY OF GOTHENBURG

# ETHICS IN THE DESIGN OF INTELLIGENT ARTIFACTS

GORDANA DODIG-CRANKOVIC

CHALMERS UNIVERSITY OF TECHNOLOGY

GOTHENBURG, SWEDEN

@CODE OF ETHICS

<https://www.ntnu.no/bigchallenge/> The Big Challenge Science Festival

The Norwegian University of Science and Technology

19th June 2019

 NTNU



# DIGITALIZATION & COGNITIZATION

## INTELLIGENCE & ARTIFICIAL INTELLIGENCE

- A symposium on the ethical implications of the digitization of our society
- Digitalization happens in parallel with introduction of cognitive properties and intelligence into artifacts
- My background: theoretical physics, computer science, philosophy of computing and ethics, computational models of cognition, recent interest: interaction design – Interdisciplinary approach



# DESIGN OF INTELLIGENT ARTIFACTS

- Ambient intelligence
- Autonomous Intelligent Systems
- Intelligent Robots & Intelligent Softbots
- Intelligent transportation
- Intelligent Cities
- Intelligent IoT
- Decision Making Algorithms



<https://bitcoinist.com/crypto-mining-becoming-concern-us-cities/>



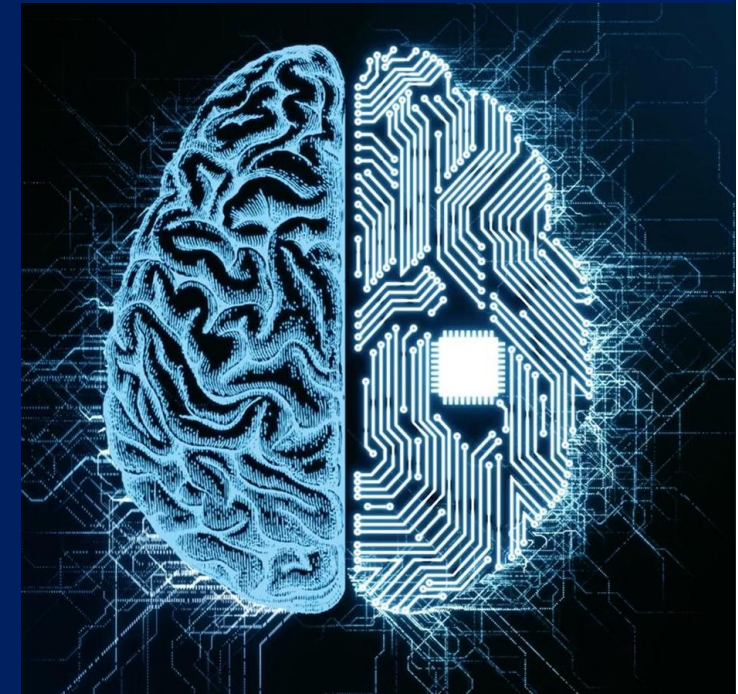
# DESIGN OF INTELLIGENT ARTIFACTS

- Intelligent Health-Care Systems
- Intelligent Personalized Medicine
- Cognitive enhancements

[Theodore Berger](#) (University of Southern California, L.A.)  
[Engineering Memories: A Cognitive Neural Prosthesis for Restoring and Enhancing Memory Function](#)

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3141091/>

<https://www.technologyreview.com/s/513681/memory-implants/>



<https://www.pbs.org/wgbh/nova/video/nova-wonders-can-we-build-a-brain/>



RECURRENT PATTERN: INSPIRATION FROM NATURE (INTELLIGENCE)

QUESTION: HOW MUCH DO WE KNOW ABOUT NATURE?

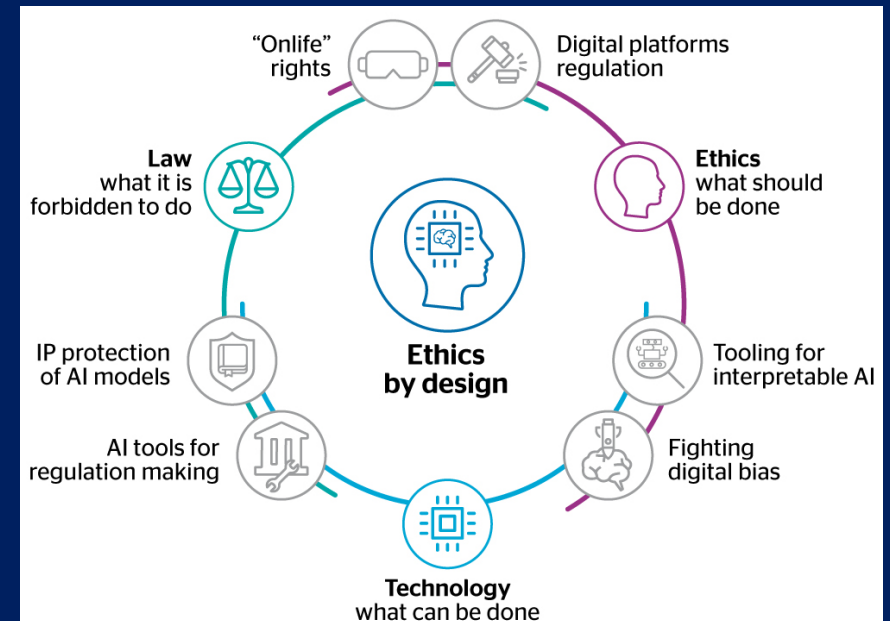


Similar to the way antibodies are able to modify the responses of blood cells to viruses, smart services will adjust their behavior according to their ecosystem's operating context.



# DO WE **WANT** INTELLIGENT EVERYTHING?

## WHY? **QUESTION OF VALUES & ETHICS**



<https://atos.net/content/mini-sites/journey-2022/human-centric-ai/>



# VALUE-SENSITIVE DESIGN

Value-sensitive design (VSD) holds that **artefacts are value-laden and design can be value-sensitive**. The approach refers to the need to identify early implicit values embedded in new technologies by focusing on the usage situations of technology.

**“Value” is defined broadly as property that a person or a group considers important in life**, and designers can intentionally inscribe their values in the design objects thus shaping them.

The design is carried out iteratively by

combining the following approaches supporting the values:

- **conceptual** (conceptions of values for users and stakeholders),
- **empirical** (how values are realized in practice)
- **technical** (design of technology),
- **research** all of which is followed by
- **assessment**



# VALUE-SENSITIVE DESIGN

## Ethical values and principles in European discussion

Expert Group/ Publication	Ethical Value/Principle	Context	Technology
Friedman et al. (2003; 2006) [1,2]	Human welfare Ownership and property Freedom from bias Universal usability Courtesy Identity Calmness Accountability (Environmental) sustainability	Value-sensitive design	ICT
Ethically Aligned Design (EAD) IEEE Global initiative (2016, 2017) [3,4]	Human benefit Responsibility Transparency Education and Awareness	Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems: Insights and recommendations for the AI/AS technologists and for IEEE standards	AI/AS



# VALUE-SENSITIVE DESIGN

## Ethical values and principles in European discussion

Asilomar AI Principles (2017) [5]	Safety Failure and juridical transparency Responsibility Values, alignment Privacy and liberty Shared benefit and prosperity Human control Non-supervision Avoiding arms race	Beneficial AI to guide the development of AI	AI
The European Group on Ethics in Science and New Technologies (EGE) (2017) [6]	Human dignity Autonomy Responsibility, Accountability Security, Safety Justice, Equality and solidarity Democracy Bodily and mental integrity Data protection and privacy Sustainability	Statement on Artificial Intelligence, Robotics and Autonomous Systems	AI, Robotics, AS



# VALUE-SENSITIVE DESIGN

## Ethical values and principles in European discussion

European Commission's High-Level Expert Group on Artificial Intelligence (AI HLEG) (2018) [7]	Respect for human dignity Freedom of the individual Respect for democracy, justice and the rule of law Equality, non-discrimination and solidarity Citizens rights Beneficence: "Do Good" Non maleficence: "Do no Harm" Autonomy: "Preserve Human Agency" Justice: "Be Fair" Explicability: "Operate transparently"	Trustworthy AI made in Europe	AI
AI4People (2018) [8]	Beneficence Non-maleficence Autonomy Justice Explicability	An ethical framework for a good AI society	AI



# VALUE-SENSITIVE DESIGN

## A preliminary set of ethical values for the context of AIS

Integrity and Human Dignity	Individuals should be respected, and AIS solutions should not violate their dignity as human beings, their rights, freedoms and cultural diversity. AIS should not threaten a user's physical or mental health.
Autonomy	Individual freedom and choice. Users should have the ability to control, cope with and make personal decisions about how to live on a day-to-day basis, according to one's own rules and preferences.
Human control	Humans should choose how or whether to delegate decisions to AIS, to accomplish human-chosen objectives.
Responsibility	Concerns the role of people and the capability of AIS to answer for the decisions and to identify errors or unexpected results. AIS should be designed so that their affects align with a plurality of fundamental human values and rights.



# VALUE-SENSITIVE DESIGN

## A preliminary set of ethical values for the context of AIS

Justice, equality, fairness and solidarity	AIS should contribute to global justice and equal access. Services should be accessible to all user groups irrespective any physical or mental deficiencies. This principle of (social) justice goes hand in hand with the principle of beneficence: AIS should benefit and empower as many people as possible.
Transparency	If an AIS causes harm, it should be possible to ascertain why. The mechanisms through which the AIS makes decisions and learns to adapt to its environment should be described, inspected and reproduced. Key decision processes should be transparent and decisions should be the result of democratic debate and public engagement.
Privacy	People should have the right to access, manage and control the data they generate.



# VALUE-SENSITIVE DESIGN

## A preliminary set of ethical values for the context of AIS

Reliability	AIS solutions should be sufficiently reliable for the purposes for which they are being used. Users need to be confident that the collected data is reliable, and that the system does not forward the data to anyone who should not have it.
Safety	Safety is an emerging property of a socio-technical system, which is created daily by decisions and activities. Safety of a system should be verified where applicable and feasible. Need to consider possible liability and insurance implications.
Security	AI should be secure in terms of malicious acts and intentional violations (unauthorized access, illegal transfer, sabotage, terrorism, etc.). Security of a system should be verified where applicable and feasible.
Accountability	Decisions and actions should be explained and justified to users and other stakeholders with whom the system interacts.



# VALUE-SENSITIVE DESIGN

## A preliminary set of ethical values for the context of AIS

Explicability	Also 'explainability'; necessary in building and maintaining citizen's trust (captures the need for accountability and transparency), and the precondition for achieving informed consent from individuals.
Sustainability	The risks of AIS being misused should be minimized: Awareness and education. "Precautionary principle": Scientific uncertainty of risk or danger should not hinder to start actions of protecting the environment or to stop usage of harmful technology.
Role of technology in society	Governance: Society should use AIS in a way that increases the quality of life and does not cause harm to anyone. Depending on what type of theory of justice a society is committed to, it may stress e.g., the principle of social justice (equality and solidarity), or the principle of autonomy (and values of individual freedom and choice).



# REFERENCES

1. Friedman, B.; Kahn, P.H., Jr. (2003) [Human values, ethics, and design](#). In The Human-Computer Interaction Handbook, Fundamentals, Evolving Technologies and Emerging Applications; Jacko, J.A., Sears, A., Eds.; Lawrence Erlbaum: Mahwah, NJ, USA; pp. 1177–1201.
2. Friedman, B.; Kahn, P.H., Jr.; Borning, A. (2006) [Value sensitive design and information systems](#). In Human-Computer Interaction in Management Information Systems: Applications; M.E. Sharpe, Inc.: New York, NY, USA; Volume 6, pp. 348–372.
3. IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems. Ethically Aligned Design, Version One – For Public Discussion (2016) [A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems](#) [https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead\\_v1.pdf](https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v1.pdf)
4. IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Ethically Aligned Design, Version 2 for Public Discussion (2017) [A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems](#). Available online: [https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead\\_v2.pdf](https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf)



# REFERENCES

5. Asilomar Conference 2017. Asilomar AI Principles. Available online: <https://futureoflife.org/ai-principles/?cn-reloaded=1>
6. European Group on Ethics in Science and New Technologies (2018) Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems. Available online: [https://ec.europa.eu/research/ege/pdf/ege\\_ai\\_statement\\_2018.pdf](https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf)
7. European Commission's High-Level Expert Group on Artificial Intelligence. Draft Ethics Guidelines for Trustworthy AI (2019) Available online: <https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai>
8. Floridi, L.; Cows, J.; Beltrametti, M.; Chatila, R.; Chazerand, P.; Dignum, V.; Luetge, C.; Madelin, R.; Pagallo, U.; Rossi, F.; et al. (2018) AI4People—An Ethical Framework for a Good AI Society. Minds Mach. 28, 689–707. <https://link.springer.com/article/10.1007%2Fs11023-018-9482-5>

All links accessed on 16 June 2019



**CHALMERS**  
UNIVERSITY OF TECHNOLOGY



UNIVERSITY OF GOTHENBURG

<https://www.ntnu.no/bigchallenge/nb/event/code-of-ethichs/>



**CHALMERS**  
UNIVERSITY OF TECHNOLOGY



UNIVERSITY OF GOTHENBURG